(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 278 659 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication of patent specification: **24.05.95** (51) Int. Cl.6: **C12N 15/05**

(21) Application number: **88300853.4**

(22) Date of filing: **02.02.88**

(54) OCS enhancer.

(30) Priority: **06.02.87 US 11614**

(43) Date of publication of application:
17.08.88 Bulletin 88/33

(45) Publication of the grant of the patent:
24.05.95 Bulletin 95/21

(84) Designated Contracting States:
AT BE CH DE ES FR GB GR IT LI LU NL SE

(56) References cited:

THE EMBO JOURNAL, vol. 6, no. 1, January 1987, pages 11-16, IRL Press Ltd,Eynsham, Oxford, GB; J.G. ELLIS et al.: "Maize Adh-1 promoter sequences controlanaerobic regulation: addition of upstream promoter elements from constitutivegenes is necessary for expression in tobacco"

THE EMBO JOURNAL, vol. 6, no. 11, November 1987, pages 3203-3208, IRL PressLtd, Oxford, GB; J.G. ELLIS et al.: "The OCS element: a 16 base pair palindrome

essential for activity of the octopine synthase enhancer"

(73) Proprietor: **LUBRIZOL ENTERPRISES, INC.**
29400 Lakeland Boulevard
Wickliffe
Ohio 44092 (US)

Proprietor: **COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION**
Limestone Avenue
Campbell,
Australian Capital Territory 2601 (AU)

(72) Inventor: **Ellis, Jeff G.**
19 Roberts Street
Macquarie
A.C.T. (AU)
Inventor: **Llewellyn, Daniel J.**
6 Myall Street
O'Connor
A.C.T. 2601 (AU)
Inventor: **Peacock, W. James**
16 Brassey Street
Deakin
A.C.T. 2600 (AU)
Inventor: **Dennis, Elizabeth**
100 Hopetown Court
Yarrahwanda
A.C.T. (AU)

EP 0 278 659 B1

UCLA SYMP. MOL. CELL. BIOL., NEW SER., vol. 62 (Plant gene systems and theirbiology), Proceedings of a CIBA-GEIGY-UCLA Symposium, Tamarron, Colorado, 2nd -8th February 1987, pages 263-277, Alan R. Liss, Inc., New York, US; J. PEACOCKet al.: "Developmental and environmental regulation of the maize alcoholdeydrogenase 1 (Adhl) gene: Promoter-enhancer interactions"

MOLECULAR AND CELLULAR BIOLOGY, vol. 7, no. 1, January 1987, pages 59-67,American Society for Microbiology; W.B. BRUCE et al.: "Functional domains of aT-DNA promoter active in crown gall tumors"

NATURE, vol. 318, no. 6046, 12th December 1985, pages 579-582, London, GB; M.P.TIMKO et al.: "Light regulation of plant gene expression by an upstreamenhancer-like element"

(74) Representative: Fisher, Adrian John et al
CARPMAELS & RANSFORD
43 Bloomsbury Square
London WC1A 2RA (GB)

Description

Field of the Invention

The field of this invention is the area of plant molecular biology and relates to plant genetic engineering by recombinant DNA technology. This invention describes the identification and characterization of a sequence of DNA from the upstream nontranscribed region of a plant-expressible gene (octopine synthase of T-DNA). This nucleotide sequence is capable of activating or increasing the transcription of nearby, preferably downstream plant-expressible genes in recombinant DNA-containing tissue from both monocotyledonous and dicotyledonous plants. The transcription activating element is useful for increasing the level of expression of a nearby gene in a plant, especially when that gene and its associated promoter are derived from heterologous plant species. Thus, the invention will facilitate the genetic engineering of plants to express novel phenotypes of economic or investigative value.

Background of the Invention

In eukaryotic genes there is a growing understanding of the DNA sequence elements which direct the initiation of transcription and which regulate or modulate gene expression. The following discussion applies to genes transcribed by RNA polymerase II. Promoters are the portions of DNA sequence at the beginnings of genes which contain the signals for RNA polymerase to begin transcription so that protein synthesis can then proceed. Eukaryotic promoters are complex, and are comprised of components which include a TATA box consensus sequence in the vicinity of about -30, and often a CAAT box consensus sequence at about -75 bp 5′ relative to the transcription start site, or cap site, which is defined as +1 (R. Breathnach and P. Chambon (1981), Ann. Rev. Biochem. 50:349-383; J. Messing et al. (1983), in Genetic Engineering of Plants, eds. T. Kosuge, C.P. Meredith, and A. Hollaender, pp. 211-227). In plants there may be substituted for the CAAT box a consensus sequence which Messing et al. (1983) have termed the AGGA box, positioned a similar distance from the cap site. Additional DNA sequences in the 5′ untranscribed region are believed to be involved in the modulation of gene expression. There are DNA sequences which affect gene expression in response to environmental stimuli, such as illumination or nutrient availability or adverse conditions including heat shock, anaerobiosis, or the presence of heavy metals. There are also DNA sequences which control gene expression during development, or in a tissue-specific fashion. Other DNA sequences have been found to elevate the overall level of expression of the nearby genes; such sequences have been termed "enhancers" in animal systems. In yeast, similar stimulatory sequences are known which are called "upstream activating sequences", which also often appear to carry regulatory information. Promoters are usually positioned 5′, or upstream, relative to the start of the coding region of the corresponding gene, and the tract containing all the ancillary elements affecting regulation or absolute levels of transcription may be comprised of less than 100 bp or as much as 1 kbp.

As defined by Khoury and Gruss (1983), Cell 33:313-314, an enhancer is one of a set of eukaryotic promoter elements that appears to increase transcriptional efficiency in a manner relatively independent of position and orientation with respect to the nearby gene. The prototype enhancer is found within the 72 bp repeat of SV40. It is located more than 100 bp upstream from the transcription start site, and has a consensus core sequence of

$$GTGG\begin{matrix}TTT\\AAA\end{matrix}G.$$

As a rule the animal or animal virus enhancers can act over a distance as much as 1 kbp 5′, in either orientation, and can act either 5′ or 3′ to the gene. The sequence motif is generally reiterated several times. Enhancers have been used in animal virus systems to study genes with weak promoters (F. Lee et al. (1981), Nature 294:228-232; A. Huang et al. (1981), Cell 27:245-255). There have been sequences from plant genes described which have homology to the animal enhancer consensus core sequence. A functional role for these sequences, however, has not been demonstrated. One example in which such homology has been found is that of the pea legumin gene 5′ in which the sequence 5′-CCACCTCC-3′ appears at about -180 relative to the transcription start site. This sequence shows about 80% homology to the complement of the animal sequence (G. Lycett et al. (1984), Nucleic Acids Res. 12:4493-4506). Two other examples where a similar sequence appears are in the 5′-flanking regions of the maize Adh1 and Adh2 genes. In those cases the sequence of note is CACCTCC, and appears at about -170 for Adh2 and -200 for Adh1 (E. Dennis et al. (1985), Nucleic Acids Res. 13:727-743; and D. Llewellyn et al. (1985), in Molecular Form and Function of the Plant Genome, eds. van Vloten-Doting, L. Groot, and T.H. Hall, Plenum Press, New York).

The yeast upstream activating sequences (UAS) have somewhat different properties than those of the animal enhancers. Like the animal

enhancers the yeast UAS's function when inserted in either orientation; they do not appear able to activate transcription when placed 3′ to the transcription start site (L. Guarente and E. Hoar (1984), Proc. Natl. Acad. Sci. USA 81:7860-7864; and K. Struhl (1984), Proc. Natl. Acad. Sci. USA 81:7865-7869). Sequences of the activating regions of some yeast promoter elements are known, and in at least two cases, homology to the SV40 enhancer consensus core sequence has been shown (B. Errede et al. (1985), Proc. Natl. Acad. Sci. USA 82:5423-5427, and G. Roeder et al. (1985), Proc. Natl. Acad. Sci. USA 82:5428-5432). Associated with these sequences is also information allowing the cell to respond to stimuli such as nutritional status or mating type, depending on the particular UAS.

Another case where upstream sequence motifs regulate downstream transcriptional activity is that of the heat shock element. It controls the response to the stress of elevated temperature in organisms from yeast to man and plants. In Drosophila the consensus sequence for the motif is 5′-CTGGAATTTCTAGA-3′ (H. Pelham and M. Bienz (1982) in Heat Shock from Bacteria to Man, Cold Spring Harbor Laboratory, pages 43-48). D. Rochester et al. (1986) EMBO J. 5:451-458, have identified two sequences 5′ to the maize hsp70 heat shock gene which are partially homologous to the consensus sequences: 5′-CCAGAGCCTTCCAGAA-3′ and 5′-CCCGAATCTTCTGGA-3′.

Recently there has been a surge of interest in plant control elements; there have been sequences proposed to be involved in tissue specificity and in responses to light and anaerobic conditions, and there have been postulated enhancer-like sequences 5′ to some highly expressed genes. One report of an enhancer-like sequence is that of J. Odell et al. (1985), Nature 313:810-812, who have described the stretch of 5′ nontranscribed region of the 35S gene of Cauliflower Mosaic Virus (CaMV) which is necessary for promoting the expression of a reporter gene. Examination of the sequence in the -105 to -46 region revealed a CAAT box-like sequence, an inverted repeat region, and a sequence resembling the animal core sequence for enhancers. It has not been demonstrated that this animal enhancer-like sequence is responsible for activity. It is known that although the host range of the CaMV is limited to members of the family Cruciferae, the entire 35S promoter does function in tobacco (J. Odell et al. (1985) supra, M. Bevan et al. (1985) EMBO J. 4:1921-1926).

Literature concerning cross expression studies, wherein a gene from one plant species is examined for expression in a different species, is growing. An early report of cross expression is that of N. Murai et al. (1983), Science 222:476-482). They reported the expression of phaseolin protein from Phaseolus vulgaris L. in sunflower (Helianthus) tissue as both a fusion protein behind a T-DNA promoter and under the control of its own promoter. Sengupta-Gopalan et al. subsequently reported that the phaseolin promoter and structural gene were functional in tobacco, and that the tissue-specific expression in the heterologous host was similar to that in the native bean host (C. Sengupta-Gopalan et al. (1985) Proc. Natl. Acad. Sci. USA 82:3320-3324).

W. Gurley et al. (1986), Mol. Cell Biol. 6:559-565; and Key et al., EPO Patent Application Serial No. 85302593.0, filed April 12, 1985, described the expression of a soybean heat shock gene in sunflower tumor tissue; the gene was strongly transcribed, and with the correct thermal induction response. Because the gene carried 3.2 kb of upstream DNA, it was presumably transcribed in response to signals carried by its own promoter.

Another example is that of J. Jones et al. (1985), EMBO J. 4:2411-2418. The promoter from a petunia chlorophyll a/b binding protein was fused to the octopine synthase gene (ocs), which provided unique sequence for detection in Northern and solution hybridization experiments. These workers found that transcription occurred in both regenerated transformed homologous (petunia) and heterologous (tobacco) plants. Ocs reporter gene activity was not detected, perhaps because the construction yielded a potential translational fusion with three amino acid substitutions at the amino terminus of the ocs polypeptide.

The first published evidence for transcription initiated from a monocot promoter in a dicot host plant comes from M. Matzke et al. (1984), EMBO J. 3:1525-1531. The maize zein Z4 gene was cloned and introduced into sunflower stemlets via a Ti-derived vector. Zein mRNA could be translated in a wheat germ system, but no zein protein was detectable in extracts of the transformed sunflower calli.

A second such report of expression across the monocot-dicot boundary is that of G. Lamppa et al. (1985), Nature 316:750-752. The wheat gene whAB1.6 encoding the major chlorophyll a/b binding protein was cloned into a T-DNA-containing vector, and transferred to both petunia and tobacco. Expression, at the level of transcription, was determined to be light-inducible and tissue-specific in the dicotyledonous hosts, as it was in the wheat. No data concerning the synthesis of the actual foreign protein were given.

D. Rochester et al. (1986), EMBO J. 5:451-458, have also detected the expression of a maize promoter in a dicotyledonous plant. The maize promoter used was that of a hybrid hsp70 gene. hsp70 is one of a set of proteins induced in maize, as in organisms from bacteria to man, in response to

heat shock. In the transgenic petunia the maize hsp70 mRNA was synthesized only in response to thermal stress.

One study of actual plant regulatory sequences is that of M. Timko et al. (1985), Nature 318:579-582. A stretch of DNA from -973 to -90 5′ to the transcriptional start site of the pea rbcS ss3.6 (ribulose 1,5 bis-phosphate carboxylase small subunit) was found to increase the level of induction of a reporter gene after illumination of transgenic tobacco plants. The stimulatory effect was observed when the -973 to -90 segment was inserted in both orientations; it did not promote high levels of gene expression when inserted 3′ to the reporter gene. J. Simpson et al. (1985), EMBO J. 4:2723-2729, studied the effect of upstream sequences from the pea chlorophyll a/b binding protein AB80 gene using an enzymatic reporter. They found that 400 bp of upstream sequence carried the necessary information for both light-induction and tissue specificity, and that sequences further upstream were involved in determining the absolute level of gene expression. In a figure showing sequence data, there is a 6 bp motif highlighted as being somewhat homologous to the animal enhancer core consensus sequence, TGGATA, which occurs at about -230 relative to the start of transcription. In neither report is there definitive data associating a specific nucleotide sequence with functional activity.

In H. Kaulen et al. (1986), EMBO J. 5:1-8, the light induction of chalcone synthase was studied using fusions of the nontranscribed region 5′ to the gene with a reporter gene. 1.2 kbp of 5′ DNA gave light inducibility and maximal expression, and deletion of the -1200 to -357 gave lower expression, but the light induction response was not reported. These authors examined the sequence and found 47 bp repeats in the region between -661 and -564; that region includes a good match to the animal enhancer consensus core sequence 5′-GTGGTTAG-3′. Enhancer activity of this sequence in plants has not been demonstrated.

There was a relatively thorough discussion of cis-active sequence involvement in light induction and tissue specificity in R. Fluhr et al. (1986), Science 232:1106-1112. They showed that the -1059 to -2 region 5′ to the pea rbcS-E9 gene gave both light inducible and tissue specific expression, and that the -352 to -2 region conferred normal expression in transgenic petunias but significantly lower levels of expression in calli. The light response was elicited only when the -37 to -2 region of 5′ DNA was present. The 5′ -410 to +15 region from the related rbcS-3A gene gave tissue specificity and light induction. In an attempt to further dissect sequence functions, they fused the -327 to -48 fragment to an enhancerless CaMV 35S promoter-reporter gene system; that fragment gave light induction and tissue specificity when inserted in both orientations. The -317 to -82 fragment from the rbcS-E9 gave similar results. Again, sequence analysis revealed regions similar to SV40 enhancers. The authors claim that these upstream stretches of DNA have the properties of light-inducible transcription enhancers; specific DNA sequences within those regions were not identified. The authors went on to discuss the analysis of seven sequenced rbcS upstream regions in which sequences similar to the SV40 enhancer core consensus and to the yeast Ty enhancer were found. These sequenced genes included representatives from Nicotiana and soybean as well as the pea. G. Morelli et al. (1985), Nature 315:200-204, reported a control sequence for dicot light-regulated genes, which is 5′-CATTATATATAGC(orA)-3′.

Two Agrobacterium tumefaciens T-DNA genes have been well characterized. The ocs gene encodes octopine synthase, and is carried on octopine-type Ti plasmids such as pTiAch5 and pTi15559. The gene for nopaline synthase is nos, and it resides on the nopaline-type Ti plasmids. Both ocs and nos and their 5′-flanking regions have been sequenced (H. DeGreve et al. (1982), J. Mol. Appl. Genet. 1:499-511; M. Bevan et al. (1983), Nucleic Acids Res. 11:369-385; A. Depicker et al. (1982), J. Mol. Appl. Genet. 1:561-573). Expression of both of these genes in plant tissue is constitutive, and there does not appear to be tissue specificity (L. Otten et al. (1981), Mol. Gen. Genet. 183:209-213).

There are no published data for enhancer-like activity in T-DNA 5′ untranscribed regions. However, C. Koncz et al. (1983), EMBO J. 2:1597-1603, did show that the region between -294 and -170 was required for full expression of ocs. The sequence for ocs was published by H. DeGreve et al. (1983), supra, after animal and animal virus enhancers were known. The authors noted the presence of a TATA box-like sequence and a polyadenylation signal at the 3′ side of the gene, but did not note any sequence of potential regulatory significance. They suggest perhaps because the ocs promoter is close to the edge of the T-DNA, there might be flanking plant sequences that influence the levels of ocs transcription.

There are conflicting data in the literature regarding the extent of 5′ sequence required for maximal expression of the nos gene. C. Koncz et al. (1983), supra, presented data that all signals required for maximal expression of the nos gene were in the 261 bp preceding the transcriptional start site. In contrast, C. Shaw et al. (1984), Nucleic Acids Res. 12:7831-7846, indicated that sequences further upstream than -88 were not essential for expression of nos in the Kalanchoe leaf and stem

test system. G. An et al. (1986), Mol. Gen. Genet. 203:245-250, have more recently established that regions of nos upstream DNA including the TATA box (-26 to -19) and the CAAT box (-78 to -70) are required for efficient transcription, and that a sequence between -130 and -101 is absolutely required for expression in tobacco. Direct and indirect repeats are revealed in that publication; deletion analysis suggests that the pair of direct repeats (-171 to -161 and -137 to 127) and the pair of inverted repeats (-148 to -141 and -114 to -106) may regulate the level of downstream gene expression.

There are several techniques available for introducing recombinant DNA into plant tissue for either stable integration into the plant genome or for measuring engineered gene activity in transient expression systems where incorporation into the genome is not required. Representative bacteria-to-plant T-DNA dependent cloning vector systems are described in G. An (1986), Plant Physiol. 81:86-91; G. An et al. (1985), EMBO J. 4:277-284; L. Herrera-Estrella et al. (1983), EMBO J. 2:987-995; L. Herrera-Estrella et al. (1983) Nature 303:209-213; and L. Herrera-Estrella et al. (1985) in Plant Genetic Engineering, ed. J.H. Dodds, New York: Cambridge University Press, pp. 63-93. The T-DNA vectors rely on mobilization from bacteria to plant using functions supplied in trans by Agrobacterium tumefaciens and its resident Ti plasmid. T-DNA mediated transfer generally is effected in such a way that stable integration into the genome results. Such systems have been primarily useful for dicotyledonous plants since the host range of Agrobacteria is thought to be limited to dicots (M. Van Montagu and J. Schell (1982) Curr. Top. Microbiol. Immunol. 96:237-254; M. DeCleene and J. Deley (1976), Bot Rev. 42:389-466) and a few noncereal monocots (J.-P. Hernalsteens et al. (1984), EMBO J. 3:3039-3041; G. Hooykaas-Van Slogteren et al., (1985), Nature 311:763-764). The most widely used plant host models for recombinant T-DNA work are the dicots sunflower, petunia, and tobacco. The technique of agroinfection has extended the range of monocots into which T-DNA-containing vectors can be introduced (N. Grimsley et al. (1986) Proc. Nat. Acad. Sci. USA 83:3282-3286).

Alternatives to the Agrobacterium-mediated DNA transfer systems are known, and include electroporation of both monocots and dicot plant protoplasts to incorporate DNA (M. Fromm et al. (1985), Proc. Natl. Acad. Sci. USA 82:5824-5828) and direct transformation of protoplasts with DNA molecules mediated by polyethylene glycol (J. Paszkowski et al. (1984), EMBO J. 3:2717-2722) or calcium ions. Another T-DNA independent means for introducing recombinant DNA is microinjection of DNA into plant cell nuclei (A. Crossway et al. (1986), Mol. Gen. Genet. 202:179-185). The techniques use plant cell protoplasts (wall-less forms) as the initial DNA recipients; known manipulations of protoplasts can result in cell or tissue culture, or ultimately in regenerated transformed plants. Use of such alternatives significantly expands the range of plants into which heterologous genes can be introduced. Paszkowski et al. (supra) have shown that integration into the genome is possible without the presence of T-DNA sequences.

The subject of this patent application is the identification of a nucleotide sequence 5′-ACGTAAGCGCTTACGT-3′ derived from the 5′-untranscribed flanking region of the ocs gene of T-DNA which activates the expression of a downstream gene driven by a plant-expressible promoter. This sequence has been termed a primary component of a plant transcription activating element. The plant transcription activating element may consist only of this component or may be a larger DNA molecule containing it, and may also comprise a second component described below. This functional primary component with enhancer-like activity in plants does not share sequence homology with the core consensus sequence of the prototype animal enhancer. Recombinant DNA constructs have been engineered with either a synthetic oligonucleotide comprising the aforementioned sequence or with the appropriate fragment of the ocs upstream region placed 5′ to the maize anaerobically-regulated alcohol dehydrogenase (Adh1) promoter with a bacterial chloramphenicol acetyl transferase (cat) reporter gene; in both instances anaerobic induction of cat enzyme activity was obtained in transformed tobacco plants. Analogous constructions without the transcriptional activating element did not give detectable expression in tobacco when either cat or Adh1 served as the reporter gene. The functionality of the transcription activating element was also determined using transient expression assays in cultured maize cells, and in cultured Nicotiana plumbaginafolia cells. Thus, the ability of the transcription activating element to function in both monocotyledonous and dicotyledonous plants has been established.

SUMMARY OF THE INVENTION

The present work describes the isolation and characterization of a transcription activating element, which acts in plants to increase or activate the expression of a downstream gene. This element is found as a naturally occurring sequence within the upstream untranscribed flanking region of the octopine synthase (ocs) gene of the T-DNA from Agrobacterium tumefaciens. The essential primary component of the plant-active transcription activating element has the identifying DNA se-

quence (5′-ACGTAAGCGCTTACGT-3). The reverse of this sequence is also effective and is also termed an identifying sequence herein. Sequences with about 50% or greater homology, and preferably about 75% or greater homology, to the identifying sequences can also function as plant transcription activating elements. Such sequences, which are necessary and sufficient for activation of downstream transcription, can be used within or as derived from their natural source. Alternatively, the sequence can be generated using known technology for chemical synthesis of DNA oligonucleotides. Also found in the natural source of the transcription activating element is a second component, with the identifying sequence 5′-GATGTTAACATC-3′. The reverse of this sequence is also effective and is also termed an identifying sequence herein. Sequences with about 50% or greater homology, and preferably 75% or greater homology to the identifying sequence will function in a similar manner. This second component is not required for plant transcription activation, but can contribute to the level of enhancement of downstream gene expression.

The principal object of this invention is to provide a transcription activating element which functions in plant tissue. This transcription activating element controls the level of transcription of the plant-expressible structural gene. It is preferably placed upstream of the transcription start site anywhere from immediately 5′ to the TATA box of the gene promoter (e.g., about -40 bp) to about 1500 bp 5′ of the transcription start site. Ideally the transcription activating element should be located between about 100 and about 300 bp 5′ to the promoter sequences, such that the level of expression of the structural gene is increased by the presence of the transcription activating element. It is preferred that the transcription activating element be placed upstream of the gene it is to regulate, but it may also be placed downstream thereof where it is effective to a lesser degree.

The transcription activating element comprises an essential primary component consisting of the sequence (5′-ACGTAAGCGCTTACGT-3′) or a sequence with about 50% to 100%, preferably about 75% to 100% homology to this sequence, which is to be used as a component of a naturally occurring sequence or to be used as a chemically synthesized entity. Such sequences are effective with the orientation given and in reverse orientation.

There can be a second, optional component to the transcription activating element, identified by the DNA sequence 5′-GATGTTAACATC-3′, or a sequence with about 50%, preferably about 75% to 100% homology to this second component. The second component preferably is placed within about 500 bp of the primary component of the transcription activating element described above, ideally within about 20 to about 100 bp of said primary component, and is effective both in the given orientation and in reverse orientation. The second component can be placed either 5′ or 3′ to the primary component, and in the given orientations, the second component is preferably 3′ to the primary component.

The transcription activation element consisting of both components is effective in reverse orientation as well as in its naturally occurring orientation with the primary component 5' to the second component. The transcription activating element functions in dicotyledonous as well as in monocotyledonous plants, but only its use in monocotyledonous plants is claimed herein.

It is a further object of this invention to provide a recombinant DNA molecule comprising a plant transcription activating element, a plant-expressible promoter, and a plant-expressible structural gene, wherein the structural gene is placed under the regulatory control of the transcription activating element and the promoter.

It is a third object of this invention to provide a method for increasing the level of expression of a plant-expressible gene in plant tissue, especially when the promoters and structural genes are derived from sources other than that plant tissue in which expression is being sought.

Brief Description of the Figures

Figure 1 gives the DNA sequence of the 176 bp fragment from the 5' untranslated region of the octopine synthase gene of the T-DNA from Agrobacterium tumefaciens. The 16 bp sequence with which the primary component of the plant-active transcription activating element is identified is underlined and labelled A. The 12 bp sequence which identifies the second component of the transcription activating element is underlined and labelled B.

Figure 2 displays the DNA sequence of the oligonucleotide carrying the primary component of the plant-active transcription activating element.

Figure 3 shows the series of pAdCAT plasmids 1-6 and the included components of the transcriptional activating element/promoter/structural gene/polyadenylation signal complex. The construction of these plasmids was achieved by recombinant DNA technology as described in the examples. Restriction sites are designated by letters: B, BamHI; E, EcoRI; Hp, HpaII, RV, EcoRV; P, PstI. Letters in brackets are sites filled in with the Klenow fragment of PolI; ocs = octopine synthase; nos nopaline synthase; cat = chloramphenicol acetyl transferase.

Nucleotide coordinates for ocs DNA are from R. Barker et al. (1983) Plant Mol. Biol. 2:335-350; and those for CaMV 355 DNA are from A. Franck et al. (1980), Cell 21:285-294.

DETAILED DESCRIPTION OF THE INVENTION

The following definitions are provided, in order to remove ambiguities as to the intent or scope of their usage in the specification and claims.

Expression refers to the transcription and translation of a structural gene contained in a recombinant DNA molecule so that a protein is made.

The term transcription activating element (TAE) refers to the DNA sequence originally discovered within the ocs 5′ untranscribed region and associated with the ability to increase the expression level of a downstream gene. The essential primary sequence component of that functional element is identified by the nucleotide sequence 5′-ACGTAAGCGCTTACGT-3′, and includes the reverse sequence thereof and sequences with about 50% or greater, and preferably about 75% or greater homology to said sequence or its reverse. It is these sequences of DNA which can, in an orientation-independent fashion, enhance (increase) or activate transcription of nearby genes in plants. The TAE is preferably located from just above the promoter TATA box to about 1500 bp, and ideally from about 50 and about 500 bp, 5′ to the transcription start site. The TAE may include as an optional second component, preferably placed within about 500 bp, and more preferably within about 20 to 100 bp of said primary component, the DNA sequence 5′-GATGTTAACATC-3′, its reverse sequence or a sequence with about 50% or greater, and preferably about 75% or greater homology to said sequence or its reverse. The TAE may be used as a chemically synthesized oligonucleotide comprising the aforementioned sequences, or a segment of naturally occurring DNA containing the sequences.

Promoter refers to sequences at the 5′-end of a structural gene which direct the initiation of transcription. Promoter sequences are necessary, but not always sufficient, to drive expression of the downstream structural genes. The promoter itself may be a composite of segments derived from more than one source, naturally occurring or synthetic. Eukaryotic promoters are commonly recognized by the presence of DNA sequences homologous to the canonical form 5′-TATAA-3′ (TATA box) about 10-30 bp 5′ to the location of the 5′-end of the mRNA (cap site, +1). About 30 bp 5′ to the TATA box another promoter component sequence is often found which is recognized by the presence of DNA sequences homologous to the canonical form 5′-CCAAT-3′ (R. Breathnach and P. Chambon (1981) Ann. Rev. Biochem. 50:349-383). In plants there may be instead a sequence known as the AGGA box, named for a symmetrical placement of adenosine residues around the triplet GNG (J. Messing et al. (1983), in Genetic Engineering of Plants, eds. T. Kosuge, C. Meredith, and A. Hollaender, Plenum Press, pp. 211-227).

Polyadenylation signal refers to any nucleic acid sequence capable of effecting mRNA processing, usually characterized by the addition of polyadenylic acid tracts to the 3′-ends of the mRNA precursors. The polyadenylation signal DNA segment may itself be a composite of segments derived from several sources, naturally occurring or synthetic, and may be from a genomic DNA or an mRNA-derived cDNA. Polyadenylation signals are commonly recognized by the presence of homology to the canonical form 5′-AATAA-3′,although variation of distance, partial "readthrough", and multiple tandem canonical sequences are not uncommon (J. Messing et al., supra). It should be recognized that a canonical "polyadenylation signal" may in fact cause transcriptional termination and not polyadenylation per se (C. Montell et al. (1983) Nature 305:600-605).

Structural gene refers to that portion of a gene comprising a DNA segment coding for a protein, polypeptide or portion thereof, possibly including a ribosome binding site and/or a translational start codon, but lacking at least one component of 5′ sequence which drives the initiation of transcription. The structural gene may be one which is not normally found in the plant cell at all or in the location at which it is introduced, in which case it is termed a heterologous structural gene. A heterologous structural gene may be derived in whole or part from any source known to the art, including a bacterial genome or episome, eukaryotic nuclear or plastid DNA, cDNA, viral DNA, or chemically synthesized DNA. It is further contemplated that a structural gene may contain one or more modifications in either the coding segments or in the untranslated regions which could affect the biological activity or the chemical structure of the expression product, the rate of expression or the manner of expression control. Such modifications include, but are not limited to, mutations, insertions, deletions, and substitutions of one or more nucleotides. The structural gene may constitute an uninterrupted coding sequence or it may include one or more introns, bounded by the appropriate plant functional splice junctions. The structural gene may be a composite of segments derived from a plurality of sources, naturally occurring or synthetic. The structural gene may also produce a fusion protein. It is contemplated that the introduction into plant tissue of recombinant DNA molecules containing the TAE/promoter/structural gene/polyadenylation sig-

nal complex will include constructions wherein the structural gene and its promoter are not derived from the same kind of plant as well as additional copies of naturally-occurring genes under TAE control.

Plant tissue includes differentiated and undifferentiated tissues of plants including, but not limited to, roots, shoots, leaves, pollen, seeds, tumor tissue, such as crown galls, and various forms of aggregations of plant cells in culture, such as embryos and calli. The plant tissue may be in planta or in organ, tissue, or cell culture.

Derived from is used herein with the meaning of "taken from" or "obtained from".

Chemically synthesized, as related to a sequence of DNA, means that the component nucleotides were assembled in vitro. Manual chemical synthesis of DNA may be accomplished using well established procedures (i.e. Caruthers (1983) in Methodology of DNA and RNA Sequencing, Weissman (ed.), Praeger Publishers (New York) Chapter 1), or automated chemical synthesis can be performed using one of a number of commercially available machines.

Regulatory control refers to the modulation of gene expression by sequence elements upstream of the transcription start site. Regulation may result in an on/off switch for transcription, or it may result in variations in the levels of gene expression. To place a gene under regulatory control of sequence elements means to place it sufficiently close to such sequence elements that the gene is switched on or off, or its level of expression is measurably varied. In this invention, the enhancer sequences are preferably placed within about 1500 bp of the structural gene and upstream therefrom.

Homology, as used herein, refers to identity of nucleotide sequences.

In summary, the identification and characterization of the transcription activating element involved the following general steps:

1. Cloning a 176 bp fragment of DNA from the upstream flanking region (-294 to -116) of the ocs gene in a construction where it was inserted upstream of the maize alcohol dehydrogenase 1 (Adh1) promoter. Expression from that promoter in a transgenic tobacco host was dependent on the presence of this ocs-derived fragment, thus demonstrating that this fragment carried transcription activating capability.
2. Demonstration that the transcription activating capability of the ocs-derived fragment of DNA was not dependent on its orientation relative to the downstream promoter sequences.
3. Analysis of the nucleotide sequence of the ocs-derived fragment to reveal two palindromic sequences: 5′-ACGTAAGCGCTTACGT-3′ and 5′-GAGTTAACATC-3′.
4. Deletion analysis of the ocs-derived fragment to determine that the presence of the 16 bp palindromic sequence was required for the activation of transcription of the downstream promoter. This established the sequence 5′-ACGTAAGCGCTTACGT-3′ as the primary component of the transcription activating element. The 12 bp palindrome (second component) could be removed without greatly reducing transcriptional activation.
5. Demonstration that the 16 bp palindromic sequence (primary component), inserted as a synthetic oligonucleotide, was sufficient to activate the transcription from a downstream promoter, and thus enhance expression of a structural gene.
6. Demonstration that the 16 bp transcription activating element could increase the expression of other promoters.
7. Demonstration that the 16 bp transcription activating element had activity in monocotyledonous as well as in dicotyledonous plant hosts.

Production of genetically modified plant tissue expressing a structural gene under control of a transcription activating element and a downstream promoter combines the specific teachings of the present disclosure with a variety of techniques and expedients known in the art. In most instances, alternative expedients exist for each stage of the overall process. The choice of expedients depends on variables such as the choice of the vector system for the introduction and stable maintenance of the transcription activating element/promoter/structural gene/polyadenylation signal combination (the expression complex), the plant species to be modified and the desired regeneration strategy, and the particular structural gene to be used, all of which present alternative process steps which those of ordinary skill are able to select and use to achieve a desired result. For instance, although the starting point for obtaining the transcription activating element is exemplified in the present application by pTiAch5, homologous DNA sequences of other octopine-type Ti plasmids, or from different sources, might be substituted as long as the appropriate modifications are made to the procedures for manipulating the DNA carrying the transcription activating element. Similarly, polyadenylation signals from the nos gene might be replaced by functional signals from other sources, again with appropriate procedural modifications. Homologs of structural genes or of other sequences may be identified by those of ordinary skill in the art by the ability of their nucleic acids to cross-hybridize under conditions of appropriate stringency as is well understood in the art. It will be understood that there may be minor sequence variations within sequences utilized or disclosed in the

present application. These variations may be determined by standard techniques to enable those of ordinary skill in the art to use the functional units of transcription activating sequence, promoter elements, structural genes, and polyadenylation signals. The use of the anaerobically induced promoter from maize Adh1 might be substituted by other DNA segments carrying promoter elements. As improved means are developed for the stable insertion of foreign genes in plant cells and for manipulating the modified cells, those of ordinary skill in the art will be able to select among those alternate process steps to achieve a desired result. Such means include, but are not limited to, electroporation, microinjection, and direct DNA transformation. These techniques expand the range of plant cells into which DNA can be introduced. The remaining steps of the preferred embodiment for obtaining a genetically modified plant include inserting the TAE/promoter/structural gene/polyadenylation signal combination into a T-DNA-containing vector, and transferring the recombinant DNA to a plant cell wherein the modified T-DNA becomes stably integrated as part of the plant genome. Techniques for in vitro culture and eventual regeneration into whole plants, which may include steps for selecting and detecting transformed plant cells, are included. Procedures for transferring the introduced gene from the originally transformed strain into commercially acceptable cultivars are known to those skilled in the art.

There are many examples of known sequences which modulate gene expression. In some cases, for example, the 6 bp core of the recognition sequence for the general amino acid control activator protein GCN4, the conservation of sequence is critical to function. In other systems, the requirements are less stringent. Precedents for divergence of sequence with maintenance of function include the heat shock elements (HSE) sequences at the 5′ ends of six Drosophila heat shock genes hsp268 and hsp27, wherein there are 7 of 14 and 9 of 14 matches to the HSE consensus sequence. For four other heat shock genes compared, homology to the consensus ranges from 11 to 13 out of 14 bases (H. Pelham and M. Bienz (1982) in Heat Shock from Bacteria to Man, eds. M. Schlesinger, M. Ashburner, and A. Tissieres, Cold Spring Harbor Laboratory, pp. 43-48). Furthermore, although the idealized sequence of the HSE is palindromic, the numbers of bases within the actual HSE's in front of those six genes which could pair in a stem-loop structure varies. Another case where there is sequence divergence while functionality is maintained is in those E. coli DNA sequences which bind the cyclic-3′,5′-adenosine monophosphate (cAMP)-cAMP receptor protein complex. The DNA binding sites have been analyzed statistically to yield a 9 bp consensus sequence (5′-AA-TGTGA--T---C-3′), found over a span of 16 bp. In six pre-gene sites studied, 8 of 9 bases matched the consensus in each case (R. Ebright (1982) in Molecular Structure and Biological Activity, eds. J. Griffin and W. Daux, New York: Elsevier Science Publishing company, pp. 91-99). In the sequence analysis of eleven half-sites to which the bacteriophage lambda repressor binds, the matches to the consensus sequence ranged from 9 of 9 to as low as 5 of 9 (T. Maniatis et al. (1975), Cell 5:109-113). It is claimed herein that a DNA sequence with about 50% or greater, and preferably about 75% or greater homology to the 16 bp sequence identified as the primary functional component within the transcription activating element will function to increase or activate the level of expression of a nearby, preferably downstream, structural gene. As will be apparent to those skilled in the art, the effectiveness of different constructions having a given percent homology to said primary component may vary, as the positions of given bases within the 16 bp identifying sequence of the primary component of the transcription activating element varies. The relative effectiveness of variant constructions can readily be ascertained without under experimentation using techniques of site-directed mutagenesis known to the art. Oligonucleotides with defined variant bases can be synthesized to substitute for the 16 bp palindromic sequence, so that positional importance of individual bases can be determined, and other known techniques can be employed for more random variations (reviewed in M. Smith (1985), Ann. Rev. Genet. 19:423-462). That about one-half of the palindromic 16 bp sequence is sufficient to activate downstream transcription is demonstratable by known molecular biological techniques, for example, by the synthesis and insertion of an appropriate 8 bp oligonucleotide.

A principal feature of the present invention in its preferred embodiment is the recombinant plasmid having an inserted heterologous promoter and heterologous structural gene whose transcriptional expression is enhanced by the transcription activating element, and in which transcription is terminated in response to the downstream polyadenylation signal. It has been determined that this transcription activating element is most effective 5′ to the promoter, and that the active sequence should be placed between about 1500 bp upstream from the transcription initiation site and immediately 5′ to the TATA sequence of the promoter, but that its orientation is not important to functionality. To be affected most strongly by the transcription activating element-promoter complex, the structural gene must be inserted on the 3′ side of said complex. (A few known promoters exert bidirectional control, in which case either side of the promoter could be

considered downstream.) That portion of the structural gene which ultimately codes for the amino terminus of the protein is the 5′-end of the gene, while that end which codes for amino acids near the carboxyl end is termed the 3′-end of the gene. The 5′-end of the structural gene is best located adjacent to the 3′-end of the TAE-promoter complex. The polyadenylation signal must be located in the correct orientation downstream from the 3′-end of the coding sequence. Another consideration is the distance between the functional elements of the expression complex. Substantial variation appears to exist with regard to these distances; therefore, the distance requirements are best described in terms of functionality. As a first approximation, reasonable operability can be obtained when the distances between functional elements are similar to those in the genes from which they were derived. An additional requirement in the case of constructions leading to fusion proteins is that the ligation of the two genes or fragments thereof must be such that the two coding sequences are in the same reading frame, a requirement well understood in the art. An exception to this requirement exists in the case where an intron separates the coding sequence derived from one gene from the coding sequence of the other. In that case, the coding sequences must be bounded by compatible splice sites, and the intron splice sites must be positioned so that the correct reading frame for both genes is established in the fusion after the introns are removed by post-transcriptional processing. Differences in rates of expression or developmental control may be observed when a given gene is inserted under the control of different transcription activating element-promoter complexes.

In the preferred embodiment the chloramphenicol acetyl transferase (cat) reporter gene has been inserted into the expression plasmid 3′ to the transcription activating element-promoter complex at a BamHI site. As will be apparent to those of ordinary skill in the art, components of the expression complex may be joined by any naturally occurring or constructed restriction sites convenient for in vitro manipulations. Incompatible ends of restriction fragments may be converted to blunt ends for ligation, or modified by the addition of linkers or adaptors. The major consideration is that the sequences at the junctions maintain transcriptional and translational functionality.

EXAMPLES

The following examples are provided for illustrative purposes only and are not intended to limit the scope of the invention.

The examples utilize many techniques well known and accessible to those skilled in the arts of molecular biology, the manipulation of recombinant DNA in plant tissue, and the culture and regeneration of transformed plants. Enzymes are obtained from commercial sources and are used according to the vendors' recommendations or other variations known in the art. Reagents, buffers and culture conditions are also known to the art. References containing standard molecular biological procedures include Maniatis et al. (1982) Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York; We (ed.) (1979) Meth. Enzymol. 68; Wu et al. (eds.) (1983) Meth. Enzymol. 100 and 101; Grossman and Moldave (eds.) Meth. Enzymol. 65; Miller (ed.) (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York; Old and Primrose (1981) Principles of Gene Manipulation, University of California Press, Berkeley; Schleif and Wensink (1982) Practical Methods in Molecular Biology; Glover (ed.) (1985) DNA Cloning Vol. I and II, IRL Press, Oxford, UK; Hames and Higgins (eds.) (1985) Nucleic Acid Hybridization, IRL Press, Oxford, UK; Sellow and Hollaender (1979) Genetic Engineering: Principles and Methods, Vols. 1-4, Plenum Press, New York, which are expressly incorporated by reference herein. Abbreviations and nomenclature, where employed, are deemed standard in the field and commonly used in professional journals such as those cited herein.

EXAMPLE 1

This example describes the cloning, transformation, and assay strategy for examining the expression of a reporter gene under the control of the maize Adh1 anaerobically inducible promoter. Use of this system in both monocot and dicot hosts such as tobacco and maize enables one to detect inserted sequences with transcriptional activating activity.

1.1 Cloning of the maize Adh1 promoter region

The maize Adh1 anaerobically inducible promoter was prepared from a genomic clone, p1S.1 (E. Dennis et al. (1984), Nucleic Acids Res. 12:3983-4000) by BamHI and HindIII digestion. The Adh1 promoter fragment contains 1200 bp 5′ and 205 bp 3′ of the ATG translation initiation codon. The fragment was ligated into pBR322 which had likewise been digested with BamHI and HindIII. The promoter fragment was shortened by cutting at a unique SacII site 11 bp 3′ of the ATG codon, deleted using the 3′-5′ exonuclease activity of T4 DNA polymerase, followed by S1 nuclease digestion to remove single-stranded ends, and repaired with the Klenow fragment of Polymerase I to ensure blunt ends. A synthetic HindIII linker was

added, the plasmids were recircularized, and several randomly chosen representatives were subcloned into M13 for sequencing. Plasmid pAd1 was deleted to +106, 2 bp upstream of the A in the translation initiation codon, and the promoter fragment extended to -1094 in the 5′ direction relative to the cap site.

1.2 Cloning of the polyadenylation signal

A plant polyadenylation signal was derived from the 3′-untranslated region of the nopaline synthase gene. The signal was obtained as a 1.7 kb EcoRI-PstI fragment of pLGV2382 (Herrera-Estrella et al. (1983), EMBO J. 2:987-995). This fragment was ligated to EcoRI and PstI-cut pAD1, the 5′ BamHI site was filled in with Klenow fragment of Polymerase I, and the HindIII site was converted to a BamHI site using synthetic linkers to yield plasmid pAd2.

1.3 Cloning of the chloramphenicol acetyl transferase gene

pNOSCAT4 (L. Herrera-Estrella, personal communication) carries the chloramphenicol acetyl transferase gene of pBR325 (F. Bolivar (1978), Gene 4:121-136). (In the construction of pNOSCAT4, the TaqI ends of the cat gene were converted to BamHI ends by the addition of linkers). pNOSCAT4 was cut with BamHI to release the cat-containing fragment which was ligated into pAd2 cut with BamHI. The resultant plasmid was cut at its unique SalI site; this site was converted to a BglII site using the appropriate linker to generate pAdcat1, in which the Adh1 promoter, the reporter gene, and the downstream polyadenylation signal are assembled in the correct orientations with respect to one another. pAdcat2 was constructed by inserting the PstI fragment containing the expression complex into PstI-cut pUC8.

An alternate source of the cat gene is Sau3A-cut pBR325; the cat gene resides on a fragment which carries 56 bp of 5′ untranslated sequence, and can be ligated with BamHI-cut pAd2. A third alternative is to isolated the appropriate TaqI fragment of pBR325, and to add BamHI linkers.

1.4 Binary vector utilization for transformation of tobacco

The binary vector pGA472 (G. An et al. (1985), EMBO J. 4:277-284) is a T-DNA-containing vector which can replicate in both E. coli and A. tumefaciens and can be mobilized between bacterial cells in triparental matings (G. Ditta et al. (1980), Proc. Natl. Acad. Sci. USA 77:7347-7351). From the Agrobacterium host it can be mobilized in trans by functions on the Ti plasmid into plants wherein it will promote stable integration into the genome via its T-DNA. pAdcat1 was inserted into pGA472 by cointegration at the unique BglII sites in each plasmid. pAdcat2 was similarly introduced by joining of the two plasmids at unique HindIII sites. These recombinant plasmids were selected in E. coli and were conjugated into A. tumefaciens strain LBA4404 Rif (A. Hoekema et al. (1983) Nature 303:209-213) for subsequent transfer into plant tissue.

Leaves of Nicotiana tabacum cv. Wisconsin 38 growing in tissue culture on Murishige and Skoog (MS) medium (T. Murishige and F. Skoog (1962), Physiol. Plant. 15:473-497) were cut into pieces (1 square cm) under liquid MS medium to prevent desiccation and added to a suspension of the transforming Agrobacterium strains. Bacteria were pregrown on YM slopes (J. Ellis et al. (1979), Physiol. Plant Pathol. 15:311-319) and then resuspended in 10 ml each MS liquid medium. After 20 minutes the infected leaf pieces were transferred to MS agar and incubated for 24h at 25 • C. The leaf pieces were then washed in sterile water and transferred to MS9 shoot induction medium (MS medium, 0.5 mg indole acetic acid, 1.0 mg benzyl aminopurine, 100 mg kanamycin sulfate, 500 mg cefotaxime per liter. When shoots were 1-2 cm high they were transferred to MS medium containing the same antibiotics. Plants that formed roots were maintained under kanamycin selection.

1.5 Assay of chloramphenicol acetyl transferase activity in transformed tobacco

The plant tissue used in the assay was either young leaves of transgenic plants or shoot cultures initiated from leaf pieces of transgenic plants on medium containing kanamycin sulfate. The plant material was made anaerobic by incubation at 28 • C on MS agar in an argon atmosphere; 18 hours was found to be sufficient for induction.

The assay was performed as described by Herrera-Estrella et al. (1983), Nature 303:209-213 except that the extraction buffer contained 0.1% cysteine-HCl and 0.1% ascorbic acid. Each µg of tissue was extracted with 1 µl of extraction buffer, and cleared by centrifugation. To 50 µl of supernatant, 0.2uCi of 14C-chloramphenicol (Amersham) was added then made to 1mM acetyl CoA and incubated at 37 • C for 30 min. The ethyl acetate extract of the reactions were concentrated by evaporation and then chromatographed on silica gel plates in chloroform-methanol (95:5). The gel plates were sprayed with fluor (0.4% PPO in 1-methylnaphthalene) and autoradiographed for 16 hours at -80 • C.

1.6 Introduction of recombinant DNA into maize protoplasts

The Zea mays c.v. Black Mexican Sweet XII-11 suspension cell line (P. Chourey and D. Zurawski (1981), Theor. Appl. Genet. 59:341-344) was cultured in modified MS medium (C. Green and R. Phillips (1975), Crop Sci. 15:417-421) at 26°C. Protoplasts were isolated according to the protocol of I. Potrykus et al. (1979), Theor. Appl. Genet. 54:209-214, and prepared for electroporation as previously described (E. Howard et al., (1987) Planta). For electroporation 100 µg of plasmid DNA was mixed with 1 ml protoplasts (3 x 10^6/ml) in HEPES-buffered saline (10mM HEPES, pH 7.2, 150mM NaCl) containing 0.2M mannitol (M. Fromm et al. (1985), Proc. Nat. Acad. Sci. USA 82:5822-5828). The cells were subjected to a 45°C heat shock for 5 min, incubated on ice 5 min, and then electroporated by a single 50 msec pulse at 250V using a capacitor-discharge electroporation apparatus, on ice. Following an additional 10 min incubation on ice, the cells were diluted tenfold with PCM (Chourey and Zurawski, supra).

1.7 Transient expression assays of reporter gene activity

Following electroporation samples of the protoplasts were divided into two aliquots; one was incubated aerobically (20% oxygen, atmospheric conditions) and the other was incubated anaerobically (5% oxygen/95% nitrogen). In both cases incubation was in the dark at 26°C for 20 hours. The cells in each aliquot were then collected by centrifugation, resuspended in 250 µl 0.25 Tris-Cl, pH 7.5, sonicated, and assayed for chloramphenicol acetyl transferase enzyme activity. Substrate and reaction products were extracted with ethyl acetate and separated by thin layer chromatography as described (C. Gorman et al. (1982), Mol. Cell. Biol. 2:1044-1051). The chromatograms were fluorographed and spots quantitated by liquid scintillation counting as previously described (Howard et al., 1987, supra).

Each plasmid construction was assayed in from two to five separate electroporation experiments. Because of the variation between protoplast preparations and the efficiency of electroporation, results were normalized to give a value of 1 for the anaerobic expression of pAdhCAT after subtraction of the nonspecific background products as measured using plasmid pAdhCAT (Howard et al., 1987. supra).

1.8 Adh1 promoter dependent expression of cat gene activity

pAdcat1 contained the maize promoter sequences found between -1094 and +106 relative to the cap site, and pAdcat2 contained that maize promoter DNA from -140 to +106. If either of these two plasmids carried cis-acting regulatory sequences which could promote anaerobic gene expression in tobacco, then cat enzyme activity should increase after a period of anaerobic induction. None of 14 plants transformed with the pAdcat1-binary vector construct expressed anaerobiosis-dependent cat activity over the background level for tobacco. Similarly, none of 29 plants transformed with the pAdcat2 construct expressed any anaerobically induced cat activity. The conclusion was that the maize Adh1 promoter activity cannot be detected in the dicot species Nicotiana tabacum using the cat reporter gene system.

When these constructs were tested in transient expression experiments in maize protoplasts, it was found that cat enzyme activity was readily detectable and anaerobically induced. Thus, it appears that the Adh1 promoter sequences are sufficient for effecting transcription initiation in a homologous system. Since there was no activity in the heterologous tobacco system, the pAdcat1 and pAdcat2 vectors might be used as probes for transcription activating elements inserted 5′ to the maize Adh1 promoter sequences.

Example 2: Activation of the maize Adh1 promoter by an upstream element from the octopine synthase gene promoter

In this example the cloning regimen which led to the discovery of the transcription-activating activity of the ocs 5′ region is elaborated.

2.1 Cloning of the transcription activating element

Ti plasmid pTiAch5 (G. de Vos et al. (1981), Plasmid 6:249-253) was digested with EcoRI and BamHI to release a 2.53 kb fragment carrying the ocs 5′ untranscribed region. The BamHI site is at position -116 with respect to the start of ocs transcription, and the fragment extends in the 5′-direction toward the EcoRI site. That fragment was inserted into EcoRI-BamHI-digested pUC8. The resulting chimera was linearized with PstI and ligated to PstI-cut pAdcat2 to generate pAdcat3. The ocs 5′-untranscribed region is inserted 5′ to the -140 to +106 portion of the maize Adh1 sequences.

2.2 Activation of maize Adh1 promoter activity

A pAdcat3 construct was introduced into tobacco cells; regenerated transformed plant tissue was prepared and assayed as described in Example 1. In at least 6 of 9 plants tested, cat enzyme activity was induced by anaerobiosis. S1 mapping experiments confirmed that transcription was initiated at the normal cap site within the Adh1 region. Therefore, it was concluded that one or more sequences found within the 5′ flanking region of the ocs gene was able to activate transcription of the maize Adh1 promoter in a tobacco host.

It is noted that there was considerable variation in the absolute levels of enzyme activity, and that some transgenic plants shown to carry the recombinant sequences by Southern hybridization experiments have no detectable activity (this work; J. Jones et al. (1986), supra). Therefore it is important to assay several transgenic plants for the desired recombinant phenotype; 20-30 are recommended for statistical validity.

Example 3: Characterization of the transcription activating element

This example describes the steps taken to localize the transcription activating element within the 5′ flanking region of the ocs gene. A 16 bp palindromic sequence 5′-ACGTAAGCGCTTACGT-3′ was identified as sufficient to activate Adh1-promoted transcription in the tobacco model system.

3.1 Localization of the transcription activating element to a 176 bp fragment

pAdcat3 was cut with BamHI and HpaII, and the 176 bp fragment, which extends from -292 to -116 relative to the start of ocs transcription was purified and ligated with pBR322 which had been digested with BamHI and ClaI to yield pOCS1. (pOCS1 can serve as a source of element-containing DNA, and it can serve as a recipient vector into which to insert promoter/foreign structural gene/polyadenylation signal complexes for subsequent manipulations.) The small EcoRI-BamHI fragment from pOCS1 was inserted into pUC8 and then the PstI fragment of pAdcat1 inserted to give pAdcat5. pAdcat6 was constructed as follows: pOCS1 DNA was cut with PstI and EcoRI and then treated with Klenow fragment and ligated to PstI linkers. After PstI digestion and removal of excess linkers, the PstI fragment of pAdcat1 was inserted. This fragment carried part of the amp gene and when inserted in the correct orientation, it complemented the deletion created in the amp gene of pOCS1 by the initial PstI-EcoRI digestion. This placed the ocs 5′ region upstream of the Adh1 promoter but in an inverted orientation with respect to pAdcat5. A HindIII linker was inserted at the SalI site of this plasmid to give pAdcat6 so that it could be ligated into the binary vector.

pAdcat5 and pAdcat6 derivatives were introduced into tobacco and those transformed plant tissues were assayed for anaerobically induced cat enzyme activity. 10 of 21 pAdcat5-transformed plants and 11 of 17 pAdcat6-transformed plants expressed inducible cat enzyme activity. Therefore the 176 bp fragment was capable of activating transcription from the maize Adh1 promoter, and this activation was independent of the orientation of that fragment relative to the Adh1 promoter region. Similar results were obtained with maize protoplasts into which these plasmids had been electroporated.

3.2 Analysis of sequence for potential regulatory regions

The published nucleotide sequence of the 5′ untranscribed region of the ocs gene was analyzed for regions of potential secondary structure which might indicate sites of regulatory function. Computer analysis revealed two regions of dyad symmetry. A 16 bp palindrome (5′-ACGTAAGCGCTTACGT-3′) was found at -194 to -179 and a 12 bp palindrome (5′-GATGTTAACATC -3′) was found in the region -149 to -138. Experiments were then designed to test these two sequences for their role(s) in transcriptional activation.

3.3 Deletion analysis of the 176 bp fragment to identify and localize the active transcriptional activating element

5′ and 3′ deletions of the 176 bp HpaII-BamHI fragment were generated using Bal31 nuclease digestion after cutting with either HpaII (for 5′ deletions) or BamHI (for 3′ deletions). After treatment with S1 nuclease, the Klenow fragment of E. coli DNA polymerase I, ligation, and transformation, the ocs-derived fragments were excised, subcloned, and sequenced as described above to determine the deletion endpoints. Suitably deleted ocs-derived fragments were then ligated to pAdhCAT-100 for functional analysis. pAdCAT-100 was prepared from pAdhCAT-140, which contains -140 to +106 of the Adh1 promoter, cat coding sequences, and the nos 3′ polyadenylation signal in pUC19. pAdhCAT-140 was cut with SmaI, digested with Bal31 and fill-in repaired with the Klenow fragment of DNA polymerase I, and modified by the addition of a SalI linker. After SalI-HindIII digestion, fragment purification after low melting temperature agarose gel electrophoresis, subcloning into pUC19 and sequencing to determine endpoints, pAdhCAT-100

was constructed by subsequent subcloning of the Adh1 region plasmid. Without the addition of a transcription activating element the pAdhCAT-100 has only background levels of reporter gene activity in either maize or tobacco protoplasts. The incorporation of the 176 bp HpaII-BamHI fragment of the ocs 5′ region restores activity comparable to that observed with pAdCAT-140 in maize protoplasts.

The series of deleted plasmids were electroporated into both maize and tobacco protoplasts and assayed in the transient expression for activation of reported gene expression. The 5′ deletions extending to -280 and -207 had full enhancer-like activity while the 5′ deletion to -168 had only a small fraction of full activity. A 5′ deletion ending at -159 had no activity. A set of 3′ deletions were also analyzed for loss of transcription activation activity. 3′ deletions extending from the BamHI site to -144, -157, and -178 all had essentially full activity. A 3′ deletion to -203 allowed no expression of the reporter gene. Thus, the sets of deletion mutants defined a region of DNA between -207 and -178 which is required for the activation of downstream gene expression. It is within this region that the 16 bp palindromic sequence lies. There appeared to be a minor transcription activating activity between -168 and the BamHI site; the 12 bp palindromic sequence is within this portion of the ocs-derived fragment.

3.4 Synthesis and activity of the synthetic 16 bp palindromic sequence

The 16 bp palindromic sequence was chemically synthesized using an automated DNA synthesizer, Model 380A (Applied Biosystems, Foster City, California). The actual sequence of the oligonucleotide which was inserted into pAdhCAT-100 is 5′-GGATCCACGTAAGCGCTTACGTGGATCC-3′. (It should be noted that the resultant palindrome extends a total of 28 bp.) The construct was electroporated into both maize and tobacco protoplasts and assayed for transcriptional activation in the transient expression system. The magnitude of the activation of cat activity was not significantly different than that observed with the HpaII-BamHI fragment derived from ocs. Similar results were obtained using or synthetic oligonucleotide with one BamHI and one EcoRI end, in a construction in which there was no extension of the palindrome. Therefore, when one considers the results from the deletion analysis and the results from the synthetic oligonucleotide, it appears that the primary component of the transcription activating element is within the 16 bp palindromic sequence 5′-ACGTAAGCGCTTACGT-3′.

3.5 Positional requirements for the transcription activating element

The 16 bp palindrome is located approximately 200 bp 5′ to the transcription start site, or about 175 bp 5′ to the TATA box in pAdcat5. The effects of increasing the distance between the transcription activating element and the cap site were tested with two novel constructs. The small BamHI-SalI fragment of pBR322 was inserted between the ocs-derived fragment and the Adh1-derived promoter fragment with the result that the distance between the palindrome and the cap site was increased to approximately 475 bp. There was an approximate loss of transcription activating ability of 25% at this distance from the signals for the initiation of transcription. A second construct contained a roughly 1200 bp fragment of Tn903 inserted between the ocs-derived and Adh1-derived fragments; thus, the intervening distance in this case was about 1400 bp. Only a small fraction of the transcription activating activity was detectable at this distance. Closer positioning of the palindromic sequence to the cap site of the Adh1 promoter was achieved by fusing the ocs-derived fragment to a series of 5′ deleted Adh1 promoter fragments. The level of expression for pAdhCAT-99 (41 bp deleted) was comparable to that for pAdhCAT5. In this case the distance between the transcription activating element and the cap site was decreased from approximately 200 bp to about 160 bp. Fusion of the ocs-derived fragment to a 5′ deletion of the Adh1 promoter to -35 yielded pAdhCAT-35, wherein the distance between the cap site and the transcription activating element was reduced to about 95 bp. Transcription activation and regulation were maintained. When the TATA box of the Adh1 promoter and the cap site were deleted, the ocs-derived fragment did not activate expression of the cat reporter gene, i.e., the ocs-derived fragment used in these experiments had no endogenous promoter activity.

Claims

1. A monocotyledonous plant transformed with a recombinant DNA molecule comprising a plant transcription activating element capable of activating or enhancing the transcription level of a gene comprising a sequence having 50% to 100% homology with an identifying sequence selected from 5'-ACGTAAGCGCTTACGT-3' and its reverse sequence.

2. A plant according to claim 1 wherein said sequence has 75% to 100% homology to said identifying sequence.

3. A plant according to claim 1 wherein said sequence has 100% homology to said identifying sequence.

4. A plant according to any of claims 1 to 3 wherein said recombinant DNA molecule also comprises a second component comprising a sequence having 50% to 100% homology with a second identifying sequence selected from 5'-GATGTTAACATC-3' and its reverse sequence.

5. A plant according to claim 4 wherein said second component has 75% to 100% homology to said second identifying sequence.

6. A plant according to claim 4 wherein said second component has 100% homology to said second identifying sequence.

7. A plant according to any preceding claim wherein said plant transcription activating element is derived from T-DNA.

8. A plant according to claim 1 wherein said transcription activating element is derived from the 5'-untranscribed region of the octopine synthase gene of T-DNA.

9. A plant according to any preceding claim comprising said recombinant DNA molecule in combination with
   (a) a plant-expressible promoter, and
   (b) a plant-expressible structural gene wherein said gene is placed under the regulatory control of said transcription activating sequence and said plant-expressible promoter.

10. A plant according to claim 9 wherein the transcription activating element is located between the 5' end of the TATA box of said promoter and about 1500 bp 5' of the transcription initiation site of said plant-expressible gene.

11. A plant according to claim 9 wherein said transcription activating element comprises a second component, consisting essentially of a sequence having 50% to 100% homology with a second identifying sequence selected from 5'-GATGTTAACATC-3' and its reverse sequence located 5' to said plant-expressible promoter, and wherein said second component is located between about 1 and about 500 bp of the identifying sequence of said DNA molecule of claim 1.

12. A plant according to any of claims 9 to 11 wherein both said elements are derived from T-DNA.

13. A plant according to claim 12 wherein both said elements are derived from an octopine synthase gene.

14. A plant according to claim 9 wherein said transcription activating element is located 3' to said promoter and said structural gene.

15. A method for enhancing the expression of a plant-expressible gene in plant tissue comprising the steps of
   (a) inserting a transcription activating element comprising a sequence having 50% to 100% homology to an identifying sequence selected from 5'-ACGTAAGCGCTTACGT-3' and its reverse sequence, in such a way that said transcription activating element modulates the expression of said gene, and
   (b) introducing said recombinant DNA molecule into plant tissue from a monocotyledonous plant.

16. The method of claim 15 wherein said recombinant DNA molecule is introduced into said plant tissue by electroporation.

17. The method of claim 15 or claim 16 wherein said transcription activating element also comprises a sequence having 50% to 100% homology to a sequence selected from 5'-GATGTTAACATC-3' and its reverse sequence.

Patentansprüche

1. Einkeimblättrige Pflanze, transformiert mit einem rekombinanten DNA-Molekül, umfassend ein die Pflanzentranskription aktivierendes Element, welches den Transkriptionspegel eines Gens aktiviert oder verstärkt, das eine Sequenz mit einer 50% bis 100% Homologie mit einer identifizierenden Sequenz, gewählt aus 5'-ACGTAAGCGCTTACGT-3' und der Umkehrsequenz davon, aufweist.

2. Pflanze nach Anspruch 1, wobei die Sequenz 75% bis 100% Homologie zur identifizierenden Sequenz aufweist.

3. Pflanze nach Anspruch 1, wobei die Sequenz 100% Homologie zur identifizierenden Sequenz aufweist.

4. Pflanze nach einem der Ansprüche 1 bis 3, wobei das rekombinante DNA-Molekül auch

eine zweite Komponente umfaßt, welche eine Sequenz mit 50% bis 100% Homologie mit einer zweiten identifizierenden Sequenz, gewählt aus 5'-GATGTTAACATC-3' und der Umkehrsequenz davon, umfaßt.

5. Pflanze nach Anspruch 4, wobei die zweite Komponente 75% bis 100% Homologie zur zweiten identifizierenden Sequenz aufweist.

6. Pflanze nach Anspruch 4, wobei die zweite Komponente 100% Homologie zur zweiten identifizierenden Sequenz aufweist.

7. Pflanze nach einem vorhergehenden Anspruch, wobei das die Pflanzentranskription aktivierende Element von T-DNA abgeleitet ist.

8. Pflanze gemäß Anspruch 1, wobei das die Transkription aktivierende Element vom 5'-untranskripierten Bereich des Octopinsynthasegens von T-DNA abgeleitet ist.

9. Pflanze nach einem vorhergehenden Anspruch, umfassend das rekombinante DNA-Molekül in Kombination mit
   (a) einem pflanzenexprimierfähigen Promotor und
   (b) einem pflanzenexprimierbaren Strukturgen, wobei das Gen unter die regulatorische Kontrolle der die Transkription aktivierenden Sequenz und des pflanzenexprimierfähigen Promotors gestellt ist.

10. Pflanze gemäß Anspruch 9, wobei das die Transkription aktivierende Element zwischem dem 5'-Ende der TATA-Box des Promotors und etwa 1500 Bp 5' der Transkriptionsinitiierungsstelle des pflanzenexprimierfähigen Gens angeordnet ist.

11. Pflanze gemäß Anspruch 9, wobei das die Transkription aktivierende Element eine zweite Komponente umfaßt, die im wesentlichen aus einer Sequenz mit 50% bis 100% Homologie mit einer zweiten identifizierenden Sequenz, gewählt aus 5'-GATGTTAACATC-3' und der Umkehrsequenz davon, angeordnet 5' zum pflanzenexprimierfähigen Promotor, umfaßt, und worin die zweite Komponente zwischen etwa 1 und etwa 500 Bp der identifizierenden Sequenz des DNA-Moleküls von Anspruch 1 angeordnet ist.

12. Pflanze nach einem der Ansprüche 9 bis 11, wobei beide Elemente von T-DNA abgeleitet sind.

13. Pflanze nach Anspruch 12, wobei beide Elemente von einem Octopinsynthasegen abgeleitet sind.

14. Pflanze nach Anspruch 9, wobei das die Transkription aktivierende Element 3' zum Promotor und zum Strukturgen angeordnet ist.

15. Verfahren zum Verstärken der Expression eines pflanzenexprimierfähigen Gens in Pflanzengewebe, umfassend die Schritte:
   (a) Inserieren eines die Transkription aktivierenden Elementes, umfassend eine Sequenz mit 50% bis 100% Homologie zu einer identifizierenden Sequenz, gewählt aus 5'-ACGTAAGCGCTTACGT-3' und der Umkehrsequenz davon, aufeine solche Weise, daß das die Transkription aktivierende Element die Expression des Gens moduliert, und
   (b) Einbringen des rekombinanten DNA-Moleküls in Pflanzengewebe einer einkeimblättrigen Pflanze.

16. Verfahren nach Anspruch 15, wobei das rekombinante DNA-Molekül in das Pflanzengewebe mittels Elektroporation eingebracht wird.

17. Verfahren nach Anspruch 15 oder Anspruch 16, wobei das die Transkription aktivierende Element auch eine Sequenz umfaßt, die 50% bis 100% Homologie zu einer Sequenz, gewählt aus 5'-GATGTTAACATC-3' und der Umkehrsequenz davon, umfaßt.

Revendications

1. Plante de la famille des monocotylédons transformée par une molécule d'ADN recombinant comprenant un élément d'activation de la transcription végétale capable d' activer ou d'amplifier le taux de transcription d'un gène comprenant une séquence présentant 50% à 100% d'homologie avec une séquence identificatrice choisie parmi 5'-ACGTAAGCGCTTACGT-3' et sa séquence inversée.

2. Plante selon la revendication 1 dans laquelle ladite séquence présente 75% à 100% d'homologie avec ladite séquence identificatrice.

3. Plante selon la revendication 1 dans laquelle ladite séquence présente 100% d'homologie avec ladite séquence identificatrice.

4. Plante selon l'une quelconque des revendications 1 à 3 dans laquelle ladite molécule d'ADN recombinant comprend également un

second composant comprenant une séquence présentant 50% à 100% d'homologie avec une seconde séquence identificatrice choisie parmi 5'-GATGTTAACATC-3' et sa séquence inversée.

5. Plante selon la revendication 4 dans laquelle ledit second composant présente 75% à 100% d'homologie avec ladite seconde séquence identificatrice.

6. Plante selon la revendication 4 dans laquelle ledit second composant présente 100% d'homologie avec ladite seconde séquence identificatrice.

7. Plante selon l'une quelconque des revendications précédentes dans laquelle ledit élément d'activation de la transcription végétale est dérivé du T-ADN.

8. Plante selon la revendication 1 dans laquelle ledit élément d'activation de la transcription est dérivé de la région 5' non transcrite du gène de l'octopine synthétase du T-ADN.

9. Plante selon l'une quelconque des revendications précédentes comprenant ladite molécule d'ADN recombinant en association avec
(a) un promoteur végétal inductible, et
(b) un gène de structure végétal inductible dans lequel ledit gène est placé sous la dépendance régulatrice de ladite séquence d' activation de la transcription et dudit promoteur végétal inductible.

10. Plante selon la revendication 9 dans laquelle l'élément d'activation de la transcription est situé entre l'extrémité 5' de la boîte TATA dudit promoteur et environ 1500 pb en 5' du site d' initiation de la transcription dudit gène végétal inductible.

11. Plante selon la revendication 9 dans laquelle ledit élément d'activation de la transcription comprend un second composant, constitué essentiellement d'une séquence présentant 50% à 100 % d'homologie avec une seconde séquence identificatrice choisie parmi 5'-GATGTTAACATC-3' et sa séquence inversée située en 5' dudit promoteur végétal inductible, et dans laquelle ledit second composant est situé entre environ 1 et environ 500 pb de la séquence identificatrice de ladite molécule d' ADN selon la revendication 1.

12. Plante selon l'une quelconque des revendications 9 à 11 dans laquelle lesdits éléments sont tous deux dérivés du T-ADN.

13. Plante selon la revendication 12 dans laquelle lesdits éléments sont tous deux dérivés du gène octopine synthétase.

14. Plante selon la revendication 9 dans laquelle ledit élément d'activation de la transcription est situé en 3' dudit promoteur et dudit gène de structure.

15. Procédé pour amplifier l'expression d'un gène végétal inductible au niveau d'un tissu végétal comprenant les étapes:
(a) d'insertion d'un élément d'activation de la transcription comprenant une séquence présentant 50% à 100% d'homologie avec une séquence identificatrice choisie parmi 5'-ACGTAAGCGCTTACGT-3' et sa séquence inversée, de telle manière que ledit élément d'activation de la transcription modifie l'expression dudit gène, et
(b) d' introduction de ladite molécule d' ADN recombinant dans un tissu végétal issu d'une plante de la famille des monocotylédons.

16. Procédé selon la revendication 15 dans lequel ladite molécule d'ADN recombinant est introduite dans ledit tissu végétal par électroporation.

17. Procédé selon la revendication 15 ou la revendication 16 dans lequel ledit élément d'activation de la transcription comprend également une séquence présentant 50% à 100% d'homologie avec une séquence choisie parmi 5'-GATGTTAACATC-3' et sa séquence inversée.

FIG. 1

The DNA Sequence of the HpaII/ BamHI Fragment from the 5'-Untranscribed Region of the ocs gene.

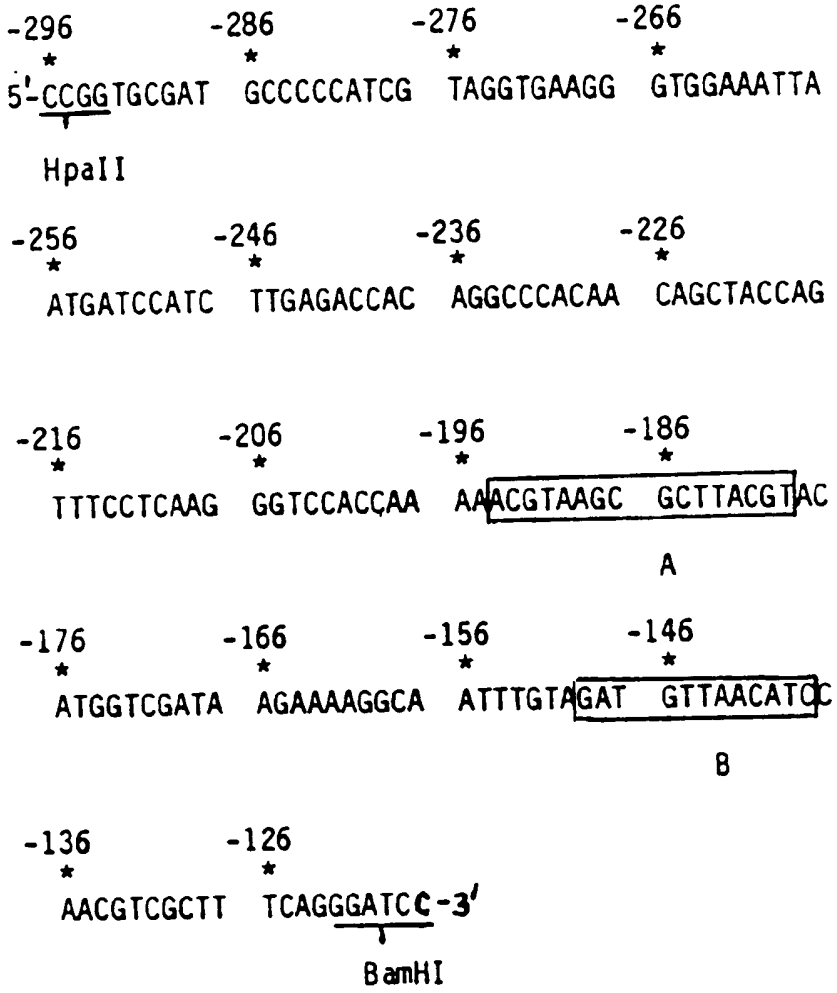

FIG. 2 The DNA sequence of the BamHI/ BamHI Synthetic Oligonucleotide Carrying the Primary Component of the Transcription Activating Element.

5'- GGATCCACGTAAGCGCTTACGTGGATCC -3'

FIG. 3